

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/104036/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Screen, Benjamin 2017. Productivity and quality when editing machine translation and translation memory outputs: an empirical analysis of English to Welsh translation. *Studia Celtica Posnaniensia* 2 (1) , pp. 113-136.  
10.1515/scp-2017-0007 file

Publishers page: <http://dx.doi.org/10.1515/scp-2017-0007>  
<<http://dx.doi.org/10.1515/scp-2017-0007>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





## PRODUCTIVITY AND QUALITY WHEN EDITING MACHINE TRANSLATION AND TRANSLATION MEMORY OUTPUTS: AN EMPIRICAL ANALYSIS OF ENGLISH TO WELSH TRANSLATION

BENJAMIN SCREEN

*Cardiff University*

### ABSTRACT

This article reports on a controlled study carried out to examine the possible benefits of editing Machine Translation and Translation Memory outputs when translating from English to Welsh. Using software capable of timing the translation process per segment, 8 professional translators each translated 75 sentences of differing match percentage, and post-edited a further 25 segments of Machine Translation. Basing the final analysis on 800 sentences and 17,440 words, the use of Fuzzy Matches in the 70-99% match range, Exact Matches and Statistical Machine Translation was found to significantly speed up the translation process. Significant correlations were also found between the processing time data of Exact Matches and Machine Translation post-editing, rather than between Fuzzy Matches and Machine Translation as expected. Two experienced translators were then asked to rate all translations for fidelity, grammaticality and style, whereby it was found that the use of translation technology either did not negatively affect translation quality compared to manual translation, or its use actually improved final quality in some cases. As well as confirming the findings of research in relation to translation technology, these findings also contradict supposed similarities between translation quality in terms of style and *post*-editing Machine Translation.

**Keywords:** Translation Memory, Machine Translation, post-editing, productivity, quality

### 1. Introduction: Machine Translation and Translation Memory

Machine Translation (MT), according to Hutchins & Somers (1992: 3), can be defined as the use of software to translate written text or oral language from one language to another, either with some assistance from humans or without it. This

assistance usually comes in the form of post-editing (or correcting the raw output to make it acceptable with a minimum of human labour (TAUS 2010). There are now a number of different systems available for Welsh, which have mostly been developed using the now dominant architecture in the field, namely Statistical Machine Translation. Whilst Machine Translation is slowly being adopted by the translation industry, this adoption goes hand in hand with the process of post-editing defined above. Translation Memory (TM), since its nascent development in the 1970s and commercialization in the 1990s, has been adopted worldwide by translators and has begun to make inroads in Wales (Watkins 2012). The basic premise of TM is that it can store previous translations of source sentences, called Exact Matches, so that the translator, and other translators using the same TM, do not have to translate the same sentence again (Somers 2003). These systems, however, also have another useful feature: they can also offer the translator target sentences that are very similar (called Fuzzy Matches), as well as reproduce the same target sentence if an identical source sentence appears again in the system. The translator can then edit these matches rather than translate ‘from scratch’, thereby saving time and cognitive effort (cf. Section 3). The similarity is calculated (usually using an algorithm called the Levenshtein Distance) and is expressed as a percentage. It is also now possible to harness the power of these two technologies by post-editing MT output within these TM systems, so that even if there is no Fuzzy or Exact match, the translator can still speed up the process by correcting the MT (Zetzsche 2007, Zaretskaya et al. 2015), or so the research reviewed below in Section 3 would lead translators to believe. Can the post-editing of MT and the use of these TM matches therefore speed up translation? Before the relevant literature is reviewed and the experimental design is described, the importance of translation in Wales and the relevance of language technology is discussed.

## 2. English to Welsh Translation in Wales

Professional translation and interpreting between Welsh and English is essential to language planning and policy in Wales, a fact that the Welsh Government has acknowledged in its recent policy document relating to Welsh (Welsh Government 2012). As Miguélez-Carballeira, Price and Kaufman (2016: 125) remind us, translation between the two languages is part of the ‘ethos’ of the brand of official bilingualism that we now see in Wales. When the British Government in the 1960s first began to respond to calls for equal treatment of Welsh and English in Wales, translation was recognized almost immediately as essential if the largely inadequate Welsh Language Act of 1967 was to be practical. As a result, the Government established the first professional Welsh

translation bureau in Cardiff in 1966 (Andrews 2015: 8). Cymdeithas Cyfieithwyr Cymru (the Welsh Association for Translators and Interpreters) was created not long after in 1976. Since then, with the Welsh Language Act 1993 and finally the Welsh Language Measure (Wales) (2011) coming into force, translation between the two languages has spiked (Kaufmann 2010, 2012). Indeed, translation in Wales, of which English to Welsh translation is a huge part, was estimated to be worth £45,000,000 per annum in a 2009 report (Prys, Prys and Jones 2009). Further developments that have raised the profile of Welsh translators and created demand for professional translation is the establishment of devolved government in Wales; the Welsh Government and the National Assembly for Wales both house large internal translation units, as do many local councils and other public sector organizations. Professional translation between Welsh and English has not only led to a spike in the public sector, however; there are also several private companies specializing in translation, interpreting and audio-visual translation between the two languages, all employing large numbers of people in language and administrative roles who themselves translate and interpret for the public and private sectors alike. This link between language planning and policy and translation then is a clear one in the case of Welsh. As González (2005: 110-111) succinctly explains, translation has an important role to play in language planning and policy, and transcends the mere facilitating of interlingual communication:

When treating translation as a mere tool to enable understanding between two parties, at least in the context of minority languages, we are clearly forgetting an essential aspect of the activity: its role in language normalisation processes (i.e. in the attempts to cause a language to be *normally* [emphasis by González] used in all spheres of a speech community) can be of equal or even more importance for such languages than the communicative function itself.

It could be considered naïve then to conceptualize translation as a service solely for the purposes of allowing one text in Language A to become understood in Language B. As a language used by 350,000 people daily in Wales (Welsh Language Commissioner and the Welsh Government 2015) and also having official status in Wales as a result of the aforementioned Welsh Language Measure (Wales) (2011), language planners and policy makers rely greatly on the translation sector to plan and provide linguistically for this speech community.

However, the prevalence of translation between Welsh and English is but one of the reasons why the relationship between productivity, quality and translation technologies was quantified for this article. The *way* translators translate, as well as *what* they translate and *why*, is also relevant. Language planners and policy makers in Wales acknowledge the role translation technology can play in facilitating professional translation work, and given the

important function translation has in language planning as elucidated above, the contribution these technologies can also make to language preservation and promotion is also widely recognized. In its previous recent language policy document, the Welsh Government (2012: 50) declared: “we must also ensure that the translation profession makes the most of the ICT tools that are available to it in order to ensure efficiency, consistency and value for money”.

The Government’s commitment to translation technology was again confirmed in a policy statement published shortly after the 2011 UK Census results for Welsh, “the reuse of translations, translation engines and automated translation for post-editing and quality control by humans, so that there can be greater prominence for Welsh” was noted as a priority (Welsh Government 2014: 11). This commitment to translation technology is just as evident in the Government’s most recent language strategy, Welsh 2050 (Welsh Government 2017). The entity responsible in Wales for policing language rights conferred on citizens by the Standards emanating from the Welsh Language Measure (Wales) (2011) is the office of the Welsh Language Commissioner. The Commissioner also publishes advice notes on best practice, and in relation to translation technology the language regulator advises that “language technology can help to accelerate and facilitate the work of the translator. It can also assist organizations who commission work to ensure the quality of translation work commissioned and to ensure value for money” (Welsh Language Commissioner 2012: 6). Given the importance of translation to official bilingualism in Wales, and the stance taken by the Welsh Government and the language regulator towards translation technology, two controlled experiments were carried out to test the apparent benefits of editing MT and TM outputs when translating from English to Welsh, in relation to productivity and final quality. Given the lack of academic research into the relationship between translation technology and Welsh despite the policies of the Welsh Government and Commissioner, this experiment is timely. Following a review of the relevant literature, these experiments are described and the results analysed.

### 3. Translation Productivity and Translation Technology

Translation technology vendors often claim that the integration of their software in translation processes, whether it be TM systems, MT systems or both, can increase translator productivity. Daelemans & Hoste (2009: 9) echo this: “commercial translation tools such as translation memories and translation workbenches are widely used and their developers claim usefulness in terms of productivity, consistency or quality. However, these claims are rarely proven using objective comparative studies”. This may indeed be true in terms of the

developers themselves but there is now a large body of academic research based on controlled and/or longitudinal studies which does lend support to the claim that editing TM and MT output as opposed to translation ‘from scratch’ (i.e. without these tools) may in fact be more efficient and does not necessarily lead to a decrease in quality (see Section 4). The claim that the translation process can be rendered more efficient by this editing process can be taken to refer to the decrease in cognitive effort as well as referring to decreases in terms of processing time and concomitant increases in words per minute/hour. Relevant studies that have investigated comparative cognitive effort between translation and post-editing MT, using pause metrics or eye-tracking variables, include O’Brien (2006b), Carl, Gutermuth and Hansen-Schirra (2015) and Koglin (2015) and those investigating the same psychological construct between translating and editing TM outputs include O’Brien (2007a), Mellinger (2014) and Screen (2016). All of these studies found that the editing process made translation cognitively easier for the translator as well as easier in terms of the physical processes of text production. Turning to translation speed and productivity, the published literature is also fairly uniform in its conclusions and includes studies by Lange and Bennett (2000), O’Brien (2006, 2007b), Offersgard et al. (2008), Brkić et al. (2009), Groves and Schmidtke (2009), Guerberof (2009, 2012, 2014), Flourney and Duran (2009), Kanavos and Kartsaklis (2010), Plitt and Masselot (2010), Skadiņš et al. (2011), Federico et al. (2012), Green et al. (2013), Aranberri et al. (2014), Elming et al. (2014), Moran et al. (2014), Silva (2014), Zhechev (2014), Uswak (2014) and Carl, Gutermuth and Hansen-Schirra (2015). As these studies informed the deductive hypotheses related to productivity measured in this study, all will now be briefly described below and the number of translators and type of source text will be given (if noted in the original publication). Only studies involving translation *professionals* have been reviewed below, but further studies that have used bilinguals or translation students when comparing translation speed are Koehn (2009), Daems et al. (2013), De Sousa et al. (2011), Lee and Liao (2011), García (2011), Yamada (2012), Vázquez et al. (2013), Läubli et al. (2013) and Depraetere et al. (2014). These also report greater efficiency when using TM and MT.

Translation between German and English is covered by Lange and Bennett’s study from 2000, although they do not identify how many translators contributed data for their results. Noting the importance of training staff to use the technology first, Lange & Bennett (2000:208) state that the use of MT and TM improved the productivity of the company’s translators significantly, and that they expected to produce over two million words using MT in the years following their investigation.

In a controlled study for her doctoral thesis, O’Brien (2006a) used a between-groups design and the key-logger Translog-II to measure cognitive,

technical (text production) and temporal (translation speed) effort. Translating a software user's guide from English to German, all 9 who post-edited the MT output finished within the allotted timeframe, but all 3 translators failed to do so. The median words per minute for the post-editors (i.e. translators who post-edited) was 17.59, but only 13.63 for the translators in the Control Group.

Offersgaard et al. (2008) carried out a small-scale test in a Danish language company, translating from English to Danish. The study concluded that the translators increased their productivity by 67% when using TM output compared to those who translated without a TM.

Brkić et al. (2009) analysed translation from English to Croatian, investigating the Déjà Vu TM system and its ability to speed up translation. Based on two professional translators translating help texts for three cameras in the system, a total of 7 minutes were saved when using TM compared with translating without it.

Groves and Schmidtke (2009) discuss how Microsoft have taken advantage of MT. The authors cite productivity gains when using MT for the following languages, with the gain expressed as a percentage in brackets: French (14.5%), Brazilian Portuguese (20%), Swedish (8%), Danish (28.6%), Czech (6.1%), Dutch (14.7%), Chinese (Simplified) (5.9%) and German (16%). The data is based on the translation of software documentation.

In a 2009 study, Guerberof (2009) carried out a productivity study based on translation from English to Spanish. All 8 translators processed 262 words of MT output, 265 words without any output and 264 words of Fuzzy Matches from the SDL Trados TM system. Her results show that 4 translators were quicker when using MT compared to using the TM or no technology, 2 using the TM output and 2 were quicker when not using any technology. Overall, the average words per minute when using MT output was 13, 12 when using the TM output and 11 without any output. For her doctoral thesis, Guerberof (2012) scaled up this study and recruited 24 professional translators. Guerberof found a significant effect of match type on translation speed, and concluded that the use of MT and TM outputs speeded up translation.

Flourney and Duran (2009) describe work done at Adobe Systems, which localizes content to over 30 languages. In two experiments, the authors concluded that productivity increased sevenfold in the first experiment based on translation from English to French, Spanish and Russian. In the second experiment, based on translation from English to Russian and Spanish, the authors calculated that translator productivity increased by 40-45%. Technical software documentation was used as a source text.

Kanavos and Kartsaklis (2010) present the results of an interesting experiment based on translation from English to Greek of two technical books. They compared the speed of translators (expressed as the number of words

produced by the hour) when using TM systems and TM and MT systems in the same workflow with the aim of translating and finally publishing those two books. By integrating the two technologies in the workflow, the authors note that “all workflows presented in this work lead to the conclusion that Machine Translation, either statistical or rule-based, has already matured and can increase substantially the productivity of professional translators” (Kanavos and Kartsaklis 2010: 20).

Plitt and Masselot (2010) investigated MT’s ability to increase productivity, using translation productivity without MT as a baseline. Translating from English to French, Italian, German and Spanish, two professional translators for each language were recruited. The data were collected over a period of two days, where the data collected on the first day was used as baseline and was then compared to the data on the second day. The number of words processed was 144,648, and software documentation was the domain of the source text. Whilst noting the dispersion in their data, Plitt and Masselot (2010) show that every translator achieved greater productivity when post-editing, with percentage gains in the region of between 20-131% depending on the participant.

Federico et al. (2012) also analyse the productivity gains that can be achieved when integrating MT to TM workflows. Basing their analysis on two experiments translating from English to German and Italian, where legal and technical documentation was translated, the authors note that productivity was increased in both experiments by an average of 27%.

Skadiņš et al. (2011) evaluate their MT system developed for translating from English to Latvian, in the context of a real localization project which included 46 tasks and 5 translators. By integrating their MT system with the SDL Trados 2009 TM system, Skadiņš et al. found that the productivity of the company’s translators increased by 32.9%, from 550 word per hour (the baseline without MT) to 731 word per hour.

Green et al. (2013) compared translation with post-editing in a controlled experiment. Sixteen translators took part in their study which was predicated on a between-groups design, with the participants translating from English to Arabic, French and German. The authors chose Wikipedia articles as the source text. Green et al. found statistically significant differences according to a two-way ANOVA between the time taken in translation and the time taken when post-editing ( $p=0.01$  for Arabic,  $p=0.01$  for French and  $p=0.001$  for German).

Aranberri et al. (2014) presents results from Spanish to Basque, and compared the potential time savings when post-editing MT compared to translation. Six translators and 6 lay users (Basque-medium university lecturers) participated in the experiment. A between groups design was used in the experiment. The source texts came from the Computer Science and Science



domain. The translators were able to improve their productivity (measured in words per hour) when post-editing text compared to those who translated without MT; a gain of 28.8% was recorded for Text A and 6.06% in the case of text B when compared to translation production speed.

As part of a project funded by the European Union to investigate cognitive processes in translation, the CASMACAT tool was developed (Elming 2014 et al.). Using this translation software, Elming et al. (2014) reports on the results of several experiments conducted in it, including measuring production speed when translating and post-editing MT. Analyzing the processing times of 5 translators translating and post-editing a set of 145 sentences in CASMACAT from English to Spanish, it was observed that every translator who participated improved productivity when post-editing, and that a productivity gain of 25% on average across all participants was achieved.

Over a period of two days, Moran et al. (2014) used a tool they developed called iOmegaT, which is capable of measuring production time per segment. Based on two translators for each language pair translating from English into German, French, Spanish and Italian, on average translator productivity increased by 54%.

Unlike many of the publications discussed above, Silva (2014) reports on the results of a longitudinal study carried out in a commercial company over a decade. Due to the complexity of measuring the productivity of individual translators over a long time period, two key performance indicators were chosen that are somewhat easier to measure, namely average lead time of projects and productivity versus lead time of projects. Silva explains how the lead time of projects over a period of ten years reduced as the use of MT in the company increased, and also sheds light on the overall increase in productivity gained gradually over the same period. When comparing the number of words produced on average in projects where MT was not used compared to projects that took advantage of MT, Silva states that an increase of 86.90% in productivity was possible when MT was integrated in to the company's workflows.

Zhechev (2014) presents the results of an experiment carried out at the Autodesk company, in which software documentation was translated. Based on translation from English to German, Spanish, French, Italian, Japanese, Korean, Polish, Brazilian Portuguese and Chinese (Simplified) and the data of four translators for each language, the experiment was carried out over a period of two days. In relation to the average baseline, productivity (measured in words per day) increased per language by the following percentages when post-editing: French (92.33%), Korean (81.93%), Italian (63.60%), Brazilian Portuguese (63.23%), Spanish (59.43%), Japanese (59.07%), Chinese (58.02%), German (44.94%) and Polish (37.13%).

Uswak (2014) contributes another study to the field from the perspective of English to German translation. Twenty professional English to German translators were recruited, and each translated 50 sentences without any support from an MT system, before translating another 50 by correcting MT output. The text was a Volkswagen car maintenance guide and 'Lucy' was the rules-based MT system used. The results showed that correcting the MT output increased the productivity of the translators (measured in words per minute), from 6.60 words per minute on average when translating, to 9.86 when correcting the output.

Finally, Carl, Gutermuth and Hansen-Schirra (2015) also investigate translation from English to German. Twelve translators took part in this experiment. Out of the twelve translators who provided data in the experiment, on average 10 out of the 12 translators were faster when post-editing compared to their translation baseline.

Following the findings of these studies, which are based on a range of TM and MT systems and an array of language pairs, as well as a range of text types, it was hypothesized that by using outputs from TM and MT engines, professional translators working from English to Welsh could also achieve greater productivity. The design of the experiment carried out to test this is described in Section 5 below. First, however, attention must be turned to quality when using translation technology.

#### 4. Translation Quality and Translation Technology

The productivity gains that are possible when using MT and TM systems are meaningless if the use of these technologies negatively affects the final quality of texts. The available literature regarding the comparative quality between texts translated with and without language technology will be reviewed below. The hypothesis for quality formulated based on this literature will then be outlined.

Fiederer and O'Brien (2009) compared translated and post-edited texts, using 11 MA students fluent in the two languages under comparison (English and German). The participants were asked to rate the translations on a scale of 1-4 for accuracy (translation accuracy, how much source text information was transferred correctly), clarity (how grammatically correct the texts were) and style. Four was the highest score. All participants were asked choose their most preferred translation also. Their results show that the post-edited texts received the highest score for accuracy and clarity, although the translations received a higher score for style. In terms of preferred translations, most participants chose the translated as opposed to the post-edited texts.

Using a group of 49 students, García (2011) compared the quality of their translated and post-edited texts using the Australian NAATI (*National Accreditation Authority for Translators and Interpreters*) guidelines. The first

professional translation reviewer who reviewed the translations using the guidelines gave a higher score to the post-edited texts in 78% of cases, and the second gave a higher mark to the same post-edited texts in 64% of cases.

Plitt and Masselot (2010) also carried out a quality assessment of their translation data reported on above. The quality assessment team rated all translation produced as either 'average' or 'good'. All texts, i.e. post-edited and translated, passed this quality assessment. The number of errors recorded was also higher in the translated texts than it was in the post-edited texts.

As part of a wider pilot study on the post-editing process, Carl et al. (2011) required 7 translation reviewers to rate four translations of the same source text according to which one they preferred (with ties being allowed), whereby two of the target texts were translated manually and the other two were post-edited. Overall, the reviewers preferred the post-edited text over the translated text, although this difference just failed to reach significance according to a Wilcoxon signed rank test ( $p = 0.0503$ ).

Skadiņš et al. (2011) analyses comparative quality following the translation and post-editing processes. A professional translation reviewer was recruited and the quality of the texts were investigated using four main parameters (accuracy [of the translation, or fidelity], terminology, style and quality of language [grammatical correctness]). Following this, the final texts were categorised according to 'excellent', 'good', 'fair', 'poor' and 'very poor'. The data showed that the introduction of MT into the translation workflow proved to be detrimental to the quality of the final texts, and that grammatical correctness was most affected.

Using a scoring system similar to that of the LISA Quality Assessment model, O'Curran (2014) asked the participating reviewers to assess the translated and post-edited texts for mistakes in Accuracy, Language, Terminology, Style, Country, Region, and Functionality (of the text, such as hidden text etc.). O'Curran (2014) found that the translated texts contained more mistakes per 1,000 words than the post-edited texts, and in terms of error type, the translated texts contained more errors related to style.

To evaluate their process data, Läubli et al. (2013) asked two experienced reviewers to review the four texts produced in their study on a scale of 1 to 4 with 4 being the highest. The reviewers rated the translated and post-edited texts according to style, grammar, syntax, semantic accuracy and quality of translation strategy. The post-edited versions of Text A and B scored 16% and 7% higher respectively than the translated versions, and 2.4% in the case of Text C. The translated version of Text D was the only translated version to be deemed better by the reviewers.

Vázquez et al. (2013) study the translation of chat forum data from English to French. Using LISA, the assessment framework also used by O'Curran (2014), they discovered that the post-edited texts received the highest score, and that the texts translated by correcting the Fuzzy Matches were of lower quality compared to post

edited texts and texts translated without either MT and TM output. However, the difference between the three types of text were not statistically significant.

The process data collated in Guerberof's 2009 study, cited above, was also subjected to a quality assessment by three experienced reviewers. Using the LISA framework, which has almost identical parameters to those used in Vázquez (2013) and O'Curran (2014), the reviewers found errors in segments that were translated by all methods (i.e. by correcting MT and TM outputs as well as manual translation). Interestingly, however, 52% of all errors were found in those segments that were produced through correcting the Fuzzy Matches, 27% in the segments that were post-edited and 21% in the segments which were manually translated. In terms of 'language' and 'mistranslation' however, 8% of errors found in the manual translations were related to 'mistranslation' but only 2% of the post-edited segments and 6% of the segments produced through correcting the Fuzzy Matches. In terms of 'language', the use of MT outputs did not seem to adversely affect this measure of quality, as 6% of errors found in post-edited segments were related to this parameter, as were 6% of errors in manually translated segments but 11% in TM corrected segments.

Using a corpus methodology, Jimenez-Crespo (2009) compared texts localized from English to Spanish, analysing the structures and content of websites and the consistency of vocabulary, syntax and typography. In order to investigate the effect of translation in TM systems on the structure and content of Spanish websites, Jimenez-Crespo (2009) analysed a corpus of 21,322 websites localized in a TM system and a corpus of 19,102 websites created in monolingual context without TM systems. He found that the structures and content of the websites localized in TM systems were more similar to English websites, but that websites created in Spanish for the Spanish-speaking community were by and large in compliance with the norms of that genre for Spanish. For example, the localized websites contained many more references to information about privacy (Data Protection, privacy, Cookies) and 'about us' sections because American websites are more likely to contain this type of information than Spanish websites as a result of the different cultural and legal background (Jimenez-Crespo 2009: 223). He also discovered that websites localized in TM systems included many more vocabulary, syntax and typographic anomalies than websites created in a monolingual Spanish context.

Daems et al. (2013) investigated translation between Dutch and English. The aim was to discover whether post-editing could lead to a final text of inferior quality compared to a translation of the same text. An original assessment framework was implemented in their experiment, based on grammatical correctness, fidelity to the original, and a number of sub-categories inside the two main categories. Two reviewers were recruited and four texts were analysed. The results showed that

post-editing resulted in better texts in the case of the first three, and that no meaningful difference in quality could be found in the case of the fourth.

Finally, Guerberof (2014) analysed the relationship between post-editing, using TM Fuzzy Matches, translation without any translation technology, and experience. The LISA quality model was used to evaluate the final texts, based on the number of errors and type. Similar to the findings of Bowker (2005) who analysed the texts of unqualified students, Guerberof (2014: 66) found that less experienced translators tend to make more errors when interacting with Fuzzy Matches than experienced translators, probably because of the lack of attention paid to the translation errors in these matches. This shows the importance of training prospective translators to use the technology correctly. What is most interesting, however, is the fact that more errors were found in the texts that were translated compared with the post-edited texts that were created by correcting the MT output (Guerberof 2014: 65-72). Results from Guerberof show that the use of MT and TM systems does not necessarily have a negative impact on the quality of the final texts, when quality is measured by calculating the number of errors found in these texts.

The results of these numerous studies then are mixed, with a majority reporting that post-editing had no detrimental impact on overall quality and several reporting how TMs can propagate errors in final texts *if the matches are not diligently reviewed*<sup>1</sup>. The working hypothesis formulated then as a result is that the use of TM and MT outputs would not lead to final texts of inferior quality compared to the texts translated without translation technology.

## 5. Data Collection

This section will describe the experimental design used, before analysing the results in Section 6.

### 5.1. Participants

In the productivity experiment, 8 professional full-time translators were recruited using convenience sampling (as were the reviewers). All were

---

<sup>1</sup> Bowker (2009), Bowker and Ehgoetz (2009) and Bowker and Ciro (2015) have also investigated comparative quality between translated texts and those texts translated by correcting the MT output, but from the perspective of recipient evaluations and the opinions of lay users. This literature review, however, is only concerned with studies that have investigated comparative quality according to linguistic evaluations by experts for reasons of space. As a result these studies will not be discussed further.

employed as professional translators of English and Welsh, all were members of a professional translation society (Cymdeithas Cyfieithwyr Cymru) and all were familiar with translation memory tools and editing Fuzzy and Exact Matches. In terms of reviewers, both had much experience of translation between Welsh and English (four was the average number of years of experience) and both had experience of proof reading translations in a professional context.

## 5.2. Experimental Design

A group of professional translators who work between English and Welsh ( $n=8$ ) were asked to translate 100 sentences from English to Welsh in a timed translation task. Time measured in seconds was the dependent variable, and the independent variable had four levels related to match type. All translators were expected to produce a translation of publishable quality. All were also aware of the origin and percentage of the match being processed as this has been shown to affect the time invested (Teixeira 2011), and so could have been a confounding variable. It is also arguably more ecologically valid to do so as all TM systems display the match percentage. In the case of Exact Matches, this was indicated by '(100%)' at the end of the target segment to be edited. No translator knew which MT system was used, however (Google Translate). SDL Trados 2015 was used to create the Fuzzy Matches, by downloading a high quality TMX file freely available to Welsh translators from the now abolished Welsh Language Board, adding and deleting content before reloading the document and using 'pretranslate' to generate the required number of matches. In terms of Exact Matches, these were left as they were found as they were already 100% matches produced by a translator previously in a real-world context. As all translators processed segments under all conditions (i.e. post-edited MT, edited Fuzzy and Exact Matches and translated manually), this experiment was predicated on a repeated measures design. A total of 17,440 words were processed; Table 1 provides further information on the corpus used. The texts to translate were taken from the public sector, and were general-purpose texts about an organization's policy on absence from work and its policy on equality in the workplace. These general-purpose texts were chosen as this is arguably the most common type of translation task for professional Welsh translators. That is, informative texts written in a more formal register and clear style from the public sector for the public at large. No time limit was set for the completion of the task and all participants were allowed to consult external linguistic resources such as dictionaries, as they would in a real-world setting. The reviewers were then asked to judge each translation on a scale of 1 to 4 for Fidelity, Accuracy and Style, as done in Fiederer and O'Brien (2009), with four being the highest (and best) score. This review was blind; no reviewer

knew who the translators were or the method through which each segment was processed (i.e. with or without MT and TM output). No time limit was set for this task either, and reviewers were paid due to the substantial amount of work involved. The amount of review work for each individual was also the reason why a simpler quality framework was utilized. Using the LISA QA framework or any other established framework in an MS Excel spreadsheet for 800 sentences would likely have been too cumbersome a task for a single reviewer.

Table 1: Details of the amount processed by each translator

Match Category	Number of Segments	Number of Source Text words	Number of Target Text words
No Match	25	481	0
Exact Match	25	329	338
Fuzzy Match	25	501	511
MT	25	378	393
<b>Total</b>	<b>100</b>	<b>2,180</b>	<b>1,754</b>

### 5.3. Apparatus

The translation task was done in an online tool used in translation process research which can record the time taken to translate per sentence in seconds. It is available from the Translation Automation Users Society (TAUS). Each translator accessed the tool online and could perform the task at home or work as per their preference. This tool allows the translator to view the previous and next segment, so that the current segment can be translated respecting its macro-context. The bilingual review was carried out in an MS Excel 2015 spreadsheet, where each set of 100 translations by the 8 translators was visible in an individual tab. The reviewer was required to select the chosen number on the 1 to 4 scale from a dropdown menu adjacent to the translations, with four the best score. Once the reviewer had finished, they were required to send the spreadsheet back via email to the author.

## 6. Results

All statistical analysis was two-tailed using a 95% confidence interval and performed in IBM SPSS.

### 6.1. Productivity Results

In order to test whether the levels of the independent variable significantly affected the dependent variable (time), a repeated measures ANOVA (Analysis of Variance) was carried out. As Mauchly's test for Sphericity was not significant ( $\chi^2 = 10.096$ ,  $df = 5$ ,  $p = .076$ ), Greenhouse-Geisser and Huynh-Feldt corrections are not reported. The ANOVA showed that match type significantly affected the amount of time invested in the translation process ( $F = 29.901$ ,  $p = .001$ ). Pairwise comparisons are reported in Table 2 below.

Table 2: Pairwise Comparisons for Match Type from the Repeated Measures ANOVA

(I) Type	(J) Type	Mean Difference (I- J)	Standard Error	Significance	95% Confidence Interval for Difference	
					Lowest	Highest
1 (No Match)	2	21108.500	6356.047	.076	-2000.864	44217.864
	3	41501.125*	5469.433	.001	21615.320	61386.930
	4	44084.375*	7332.126	.003	17426.175	70742.575
2 (Fuzzy Match)	1	-21108.500	6356.047	.076	-44217.864	2000.864
	3	20392.625*	4672.389	.020	3404.720	37380.530
	4	22975.875*	3700.738	.003	9520.703	36431.047
3 (MT)	1	-41501.125*	5469.433	.001	-61386.930	-21615.320
	2	-20392.625*	4672.389	.020	-37380.530	-3404.720
	4	2583.250	3134.737	1.000	-8814.049	13980.549
4 (Exact Match)	1	-44084.375*	7332.126	.003	-70742.575	-17426.175
	2	-22975.875*	3700.738	.003	-36431.047	-9520.703
	3	-2583.250	3134.737	1.000	-13980.549	8814.049

From this it can be concluded that statistically significant differences are to be found between No Match (or manual translation) and post-editing MT ( $p = .001$ ), No Match and Exact Matches ( $p = .003$ ), but not between No Match and Fuzzy Match in the 70-99% range ( $p = .076$ ). However, as can be seen in Table 3 below, all translators except Translator 8 made significant time gains when editing these Fuzzy Matches. The fact that Translator 8 did not could be the



reason why the difference overall just failed to reach statistical significance. The use of MT (Google Translate) and TM (SDL Trados 2015) outputs in this controlled task therefore significantly speeded up translation. The temporal difference between the levels of the independent variable can be seen clearly in Figure 1 below, where seconds is the scale of measurement on the y-axis. Figure 1 is a boxplot, or box and whisker diagram. The black line in the middle denotes the central tendency, in this case the median which is less influenced by extreme values, and the white box displays 50% of the data. The top line shows the upper 25% of the data the bottom line the lower 25%.

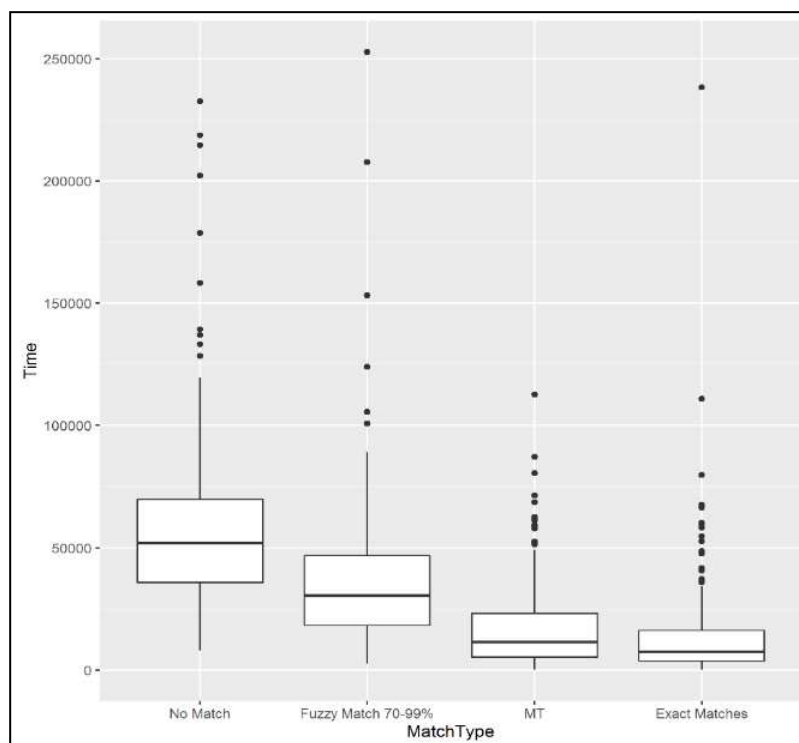


Figure 1: Match Type and Time (seconds)

Table 3: Time savings as a percentage

Participant	Match Type	No Match	Saving %
1	Fuzzy Match	14	27.27%
	Exact Match		366.67%
	MT		180%
2	Fuzzy Match	29	123.08%
	Exact Match		866.67%
	MT		480%
3	Fuzzy Match	22	83.33%
	Exact Match		633.33%
	MT		340%
4	Fuzzy Match	23	43.75%
	Exact Match		91.67%
	MT		130%
5	Fuzzy Match	31	6.90%
	Exact Match		210%
	MT		287.50%
6	Fuzzy Match	34	161.54%
	Exact Match		1600%
	MT		209.09%
7	Fuzzy Match	27	80%
	Exact Match		350%
	MT		285.71%
8	Fuzzy Match	12	-0.1
	Exact Match		100%
	MT		200%

Table 3 above details the temporal gain achieved when working with TM and MT outputs, expressed as a percentage per translator and using their No Match ('from scratch') translation time in minutes as a baseline.

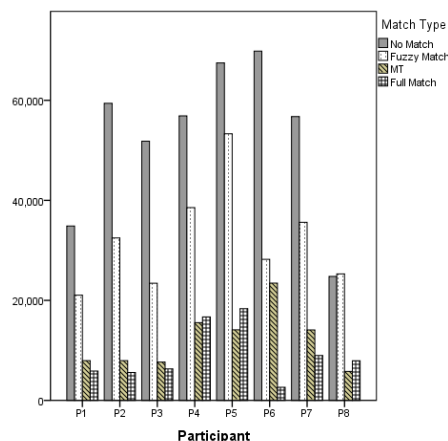


Figure 2: Median Processing Time in seconds per Welsh translator

There was also a statistically significant correlation between all translators' Exact Match time data and their MT data ( $t = .098$ ,  $n = 200$ ,  $p = .041$ ), according to Kendall's tau-b non-parametric bivariate correlation test, but not between the Fuzzy Match and MT data ( $t = 0.83$ ,  $n = 200$ ,  $p = .081$ ). Given that O'Brien (2006b) found a correlation between the cognitive effort exerted in Fuzzy Match (80-99%) editing and MT post-editing, this is an interesting finding (but naturally limited to the system and language pair here investigated). Figure 2 charts the central tendency of translation time (or processing time) in seconds per translator. As can be seen, every translator improved upon their time when using either MT or matches, and that in 50% of cases MT required the least amount of time to be invested in translation, and Exact Matches in the other 50% of cases.

Finally, mean words per minute (WPM) for each match type was calculated and was as follows: mean WPM for the No Match category was 23, 36 for the Fuzzy Matches in the 70-99% range, 62 for Exact Matches and an average WPM of 83 was recorded for sentences that were post-edited.

## 6.2. Quality Results

The quality results will now be analysed. Wilcoxon Signed Rank tests were used to compare each individual set of scores with their No Match score counterparts. This test was chosen due to the use of ordinal data and because related samples were being compared. Table 4 below details the  $p$ -value for each test.

Table 4: Results of Wilcoxon Signed Rank Tests

	<b>Style</b>	<b>Grammatical</b>	<b>Fidelity</b>
Fuzzy Match v. No	$p = .657$	$p = .837$	$p = .281$
Exact match v. No	$p = .004$	$p = .000$	$p = .068$
Post-editing v. No	$p = .200$	$p = .054$	$p = .000$

Firstly, when the Fuzzy and No Match results are compared, no statistically significant differences were found for either category. The use of Fuzzy Matches then did not lead to final texts that were less aesthetic stylistically or contained more grammar or fidelity errors. The results for Exact Matches, however, change this picture. The results show the use of Exact Matches lead to a statistically significant *increase* in the scores awarded by the reviewers for style (Exact Match mean = 3.75, No Match mean = 3.65) and for grammatical accuracy (Exact Match mean = 3.83, No Match mean = 3.70). No such difference was found, however, for fidelity. Finally, post-editing MT did not lead to final texts that contained more grammar and stylistic errors compared to manual

translation (No Match), as no statistically significant differences were found. It was found, however, that post-editing MT in this study *decreased* the likelihood of mistranslation, as there was a statistically significant difference between the fidelity scores (Post-editing mean = 3.89, No Match mean = 3.74). Figure 3 displays the mean for each category per match type.

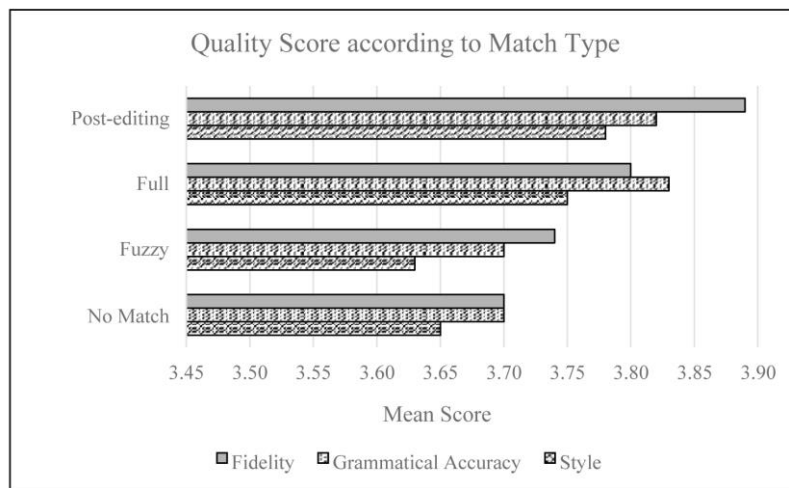


Figure 3: Quality Score according to Match Type

The translations upon which the productivity results were based therefore tended to be of a high quality, thereby validating those results.

## 7. Discussion and Conclusion

The deductive hypotheses based on the large number of studies reviewed above, that the use of Fuzzy and Exact Matches as well as post-editing MT could significantly speed up translation between English and Welsh whilst not negatively affecting quality, can be accepted. Statistically significant results were found between manual translation time and translation time when correcting TM output and raw MT output provided by Google Translate. Using TM and MT output allowed the professional translators to achieve significant productivity gains compared to their translation time when not using this output. The quality results show this increase in productivity does not lead to a decrease in overall quality, and that in fact in some cases quality can either be as high as when manually translating, or that it can even be improved. These results

confirm the findings of others in terms of productivity reviewed in Section 3. In terms of quality, Fiederer and O'Brien (2009) found that manually translated segments were rated better for quality ( $p = .0001$ ) using the same statistical test; this result was not repeated here. Guerberof (2009) found that most errors were found in the translations produced through editing Fuzzy Matches, again this result was not repeated either. Skadiņš et al. (2011) found that MT led to a decrease in overall quality following an increase in productivity; this quality result was also not found in the current study. The use of translation technology therefore can indeed be beneficial to translators working from English to Welsh, as it is to translators of other languages. Given the importance of the translation sector to language planning in Wales, as indeed it is to Irish in the Republic of Ireland (Ní Ghearáin 2007, O'Connell and Walsh 2006) and Northern Ireland (Núñez 2013), these results are important. This is especially true given the low take-up of these technologies by the Welsh translation community thus far (Welsh Language Commissioner 2015). One limitation of this study is that only one directionality was investigated (English to Welsh); it would now be beneficial to investigate Welsh to English translation also. It should be borne in mind, however, that most professional translation in Wales is carried out in the directionality here investigated. A further avenue for research based on the current data set would be to analyse the specific types of errors made and investigate whether the use of translation technology led to certain kinds of errors, and how the errors made in English to Welsh translation differ from the types of mistakes made in the translation of other languages.

## REFERENCES

- Andrews, Tegau. 2015. *Cyd-destun gwleidyddol a chymdeithasol cyfieithu yn y Gymru gyfoes*. Available at: <https://www.cadarn.ac.uk/cy/adnodd/ysgrifau-chanllawiau-cyfieithu-%E2%80%93-delyth-prys-robot-trefor-goln> (accessed: 27 January 2016).
- Aranberri, Nora et al. 2014. "Comparison of post-editing productivity between professional translators and lay users", in: Sharon O'Brien, Michel Simard and Lucia Specia (eds.) *Proceedings of the Third Workshop on Post-Editing Technology and Practice*. Vancouver, Canada, October 2014: AMTA, 20-33.
- Bowker, Lynne and Jairo Ciro. 2015. "Investigating the usefulness of machine translation for newcomers at the library", *Translation and Interpreting Studies*, 10(2): 165-186.
- Bowker, Lynne and Melissa Ehgoetz 2009. "Exploring user acceptance of Machine Translation output", in: Dorothy Kenny and Ryou Kyongjoo (eds.) *Across Boundries: International Perspectives on Translation Studies*. Newcastle: Cambridge Scholars Publishing, 209-225.
- Bowker, Lynne. 2005. "Productivity v. Quality? A pilot study on the impact of translation memory systems", *Localisation Focus* 15: 3-20.

- Bowker, Lynne. 2009. "Can Machine Translation meet the needs of official language minority communities in Canada? A recipient evaluation". *Linguistica Antverpiensia* 8: 123-155.
- Brkić, Marija et al. 2009. *Using Translation Memory to speed up translation process*. Available at: <http://darhiv.ffzg.unizg.hr/8036/> (accessed 14 August 2016).
- Carl, Michael et al. 2011. *The process of post-editing: A pilot study*. Available at: <http://mt-archive.info/NLPSC-2011-Carl-1.pdf> (accessed: 11 March 2016).
- Carl, Michael et al. 2015. "Post-editing Machine Translation: Efficiency, strategies and revision processes in professional translation settings", in: Aline Ferreira and John W. Schwieter (eds.) *Psycholinguistic and cognitive inquiries into translation and interpreting*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 145-174.
- Daelemans, Walter and Veronique Hoste. 2009. "Introduction. Evaluation of translation technology", *Linguistica Antverpiensia* 8: 9-17.
- Daems, Joke, Lieve Macken and Sonia Vandepitte. 2013. "Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE", in: Sharon O'Brien et al. (eds.) *Proceedings of the MT Summit XIV Workshop on Post-editing Technology and Practice*. Nice, France, September 2 2013: AMTA, 63-71.
- De Sousa, Sheila et al. 2011. "Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles", in: Galia Angelova et al. (eds.) *Proceedings of the International Conference in Recent Advances in Natural Language Processing*. Hissar, Bulgaria, 12-14 September, 2011: INCOMA, 97-104.
- Depraetere, Ilse et al. 2014. "Post-Edited quality, post-editing behaviour and human evaluation: A case study", in: Sharon O'Brien et al. (eds.) *Post-editing of Machine Translation: Processes and applications*. Newcastle: Cambridge Scholars Publishing, 78-109.
- Elming, Jakob et al. 2014. "Investigating user behaviour in post-editing and translation using the CASMACAT workbench", in: Sharon O'Brien et al. (eds.) *Post-editing Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, 147-169.
- Federico, Marcello et al. 2012. *Measuring user productivity in Machine Translation enhanced computer assisted translation*. Available at: [https://www.matecat.com/wp-content/uploads/2012/11/Measuring\\_User\\_Productivity\\_in\\_Machine\\_Translation\\_Enhanced\\_matecat.pdf](https://www.matecat.com/wp-content/uploads/2012/11/Measuring_User_Productivity_in_Machine_Translation_Enhanced_matecat.pdf) (accessed 07 March /2016).
- Fiederer, Rebecca. and Sharon O'Brien. 2009. "Quality and Machine Translation: A realistic objective?", *Journal of Specialised Translation* 11: 52-74.
- Flourney, Raymond and Christine Duran. 2009. "Machine Translation and document localization at Adobe: From pilot to production", in: Isabelle Pierre et al. (eds.) *Proceedings of the 12th MT Summit*. Ottawa, August 26-30, 2009: AMTA.
- García, Ignacio. 2011. "Translating by post-editing: Is it the way forward?" *Machine Translation* 25(3): 217-237.
- González, Marta García. 2005. "Translation of minority languages in bilingual and multilingual communities", in: Albert Branchadell and Lovell Margaret West (eds.) *Less Translated Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 105-125.
- Goves, Declan and Schmidtke, Dag. 2009. "Identification and analysis of post-editing patterns for MT", in: Isabelle Pierre et al. (eds.) *Proceedings of the 12th MT Summit*. Ottawa, August 26-30, 2009: AMTA.

- Green, Spence, Jeffrey Heer and Christopher Manning. 2013. *The efficacy of human post-Editing for language translation*. Available at: <https://nlp.stanford.edu/pubs/green+heer+manning.chi13.pdf> (accessed 13 May 2015).
- Guerberof, Ana Arenas. 2009. "Productivity and quality in the post-editing of outputs from translation memories and Machine Translation", *International Journal of Localization* 7(1): 11-21.
- Guerberof, Ana Arenas. 2012. *Productivity and quality in the post-editing of outputs from Translation Memories and Machine Translation*. Unpublished PhD Thesis, Universitat Rovira I Virgili.
- Guerberof, Ana Arenas. 2014. "The role of professional experience in post-editing from a quality and productivity perspective", in: Sharon O'Brien et al. (eds.) *Post-editing of Machine Translation: Processes and applications*. Newcastle: Cambridge Scholars Publishing, 51-76.
- Hutchins, John and Harold Somers. 1992. *Introduction to Machine Translation*. London: Academic Press.
- Jimenez-Crespo, Miguel. 2009. "The effect of Translation Memory tools in translated web texts: Evidence from a comparative product-based study", *Linguistica Antverpiensia* 8: 213-232.
- Kanavos, Panagiotis and Dimitrios Kartsaklis, 2010. "Integrating Machine Translation with Translation Memory: A practical approach", in: Ventsislav Zhechev (eds.) *Proceedings of the second joint EM+/CNGL workshop "Bringing MT to the user: Research on integrating MT in the translation industry"*. Denver, Colorado 4 Novemebr, 2010: AMTA, 11-20.
- Kaufmann, Judith. 2010. "Cyfieithu a pholisi iaith", *Contemporary Wales* 23(1): 171-183.
- Kaufmann, Judith. 2012. "The Darkened glass of bilingualism? Translation and interpreting in Welsh language planning", *Translation Studies* 5(3): 327-344.
- Koehn, Phillip. 2009. "A process study of computer-aided translation", *Machine Translation* 23(4): 241-263.
- Koglin, Arlene. 2015. "An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors", *Translation and Interpreting* 7(1): 126-14
- Lange, Carmen Andres and Winfield Scott Bennett. 2000. "Combining Machine Translation with Translation Memory at Baan", in: Robert C. Sprung (ed.) *Translating into Success: Cutting-edge strategies for going multilingual in a global age*. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 203-219.
- Läubli, Samuel et al. 2013. "Assessing post-editing efficiency in a realistic translation environment", in: Sharon O'Brien, Michel Simard and Lucia Specia (eds.) *Proceedings of the XIV MT summit on post-editing technology and practice*. Nice, France, 6-9 September 2013: EAMT, 83-91.
- Lee, Jason and Posen Liao. 2011. "A comparative study of human translation and Machine Translation with post-editing", *Compilation and Translation Review* 4(2): 104-149.
- Mellinger, Christopher Davey. 2014. *Computer-Assisted translation: An investigation of cognitive effort*. Unpublished PhD Thesis, Kent State University.
- Miguélez-Carballeira, Helena Angharad Price and Judith Kaufman et al. 2016. "Introduction: Translation in Wales: History, theory and approaches", *Translation Studies* 9(2): 125-136.
- Moran, John, David Lewis and Christian Saam. 2014. "Analysis of post-editing data: A productivity field test using an instrumented CAT tool", in: Sharon O'Brien et al.

- (eds.) *Post-editing of Machine Translation: Processes and applications*. Newcastle: Cambridge Scholars Publishing, 126-147.
- Ní Ghearáin, Helena. 2007. "Yr iaith Wyddeleg a'r Gyfraith: Rhai materion o ran statws a chorpws", *Cambrian Law Review*, 38: 1-38.
- Núñez, Gabriel González. 2013. "Translating for linguistic minorities in Northern Ireland: A look at translation policy in the judiciary, healthcare and local government", *Current Issues in Language Planning* 14(3-4): 474-489.
- O'Brien, Sharon. 2006. "Pauses as indicators of cognitive effort in post-editing Machine Translation output", *Across Languages and Cultures* 7(1), 1-21.
- O'Brien, Sharon. 2007a. "An empirical investigation of temporal and technical post-editing effort", *Translation and Interpreting Studies* 2(1): 83-136.
- O'Brien, Sharon. 2007b. "Eye-tracking and Translation Memory matches", *Perspectives: Studies in Translatology* 14(3), 185-205.
- O'Connell, Eithne and John Walsh. 2006. "The translation boom: Irish and language planning in the twenty-first century", *Administration*, 54(3): 22-43.
- O'Curran, Elaine. 2014. Machine Translation and post-editing for user generated content: An LSP perspective. Available at: [http://www.amtaweb.org/AMTA2014Proceedings/AMTA-2014Proceedings\\_Research Track\\_final.pdf](http://www.amtaweb.org/AMTA2014Proceedings/AMTA-2014Proceedings_Research Track_final.pdf) (accessed: 12 May 2015).
- Offersgaard, Lene et al. 2008. *Domain specific MT in use*. Available at: <http://www.mt-archive.info/05/EAMT-2008-Offersgaard.pdf> (accessed: 4 September 2015).
- Plitt, Mirko and François Masselot. 2010. "A productivity test of Statistical Machine Translation post-editing in a typical localisation context", *The Prague Bulletin of Mathematical Linguistics* 93: 7-16.
- Prys, Delyth, Gruffudd Prys and Dewi Bryn Jones. 2009. *Gwell Offer Technoleg Cyfieithu ar gyfer y Diwydiant Cyfieithu yng Nghymru: Arolwg Dadansoddol*. Available at: [http://techiaith.bangor.ac.uk/wp-content/uploads/2016/06/CATcymru\\_Adroddiad-Terfynol\\_HE06fsp\\_CY.pdf](http://techiaith.bangor.ac.uk/wp-content/uploads/2016/06/CATcymru_Adroddiad-Terfynol_HE06fsp_CY.pdf) (accessed: 15 May 2015).
- Screen, Benjamin. 2016. "What does Translation Memory do to translation? The effect of Translation Memory output on specific aspects of the translation process", *Translation & Interpreting* 8(1): 1-17.
- Silva, Roberto. 2014. "Integrating post-editing MT in a professional translation workflow", in: Sharon O'Brien et al. (eds.) *Post-editing Machine Translation: Processes and applications*. Newcastle: Cambridge Scholars Publishing, 24-51.
- Skadiņš, Raivis et al. 2011. *Evaluation of SMT in localization to under-resourced inflected language*. Available: <http://mt-archive.info/EAMT-2011-Skadins.pdf> (accessed: 26 September 2016).
- Somers, Harold. 2003. "Translation memory systems", in: Harold Somers (ed.) *Computers and translation: A Translator's guide*. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 31-49.
- TAUS. 2010. *Post-editing in practice*. Available at: <https://www.taus.net/think-tank/reports/postedit-reports/postediting-in-practice> (accessed 07 December 2015).
- Teixeira, Carlos. 2011. "Knowledge of provenance and its effects on translation performance in an integrated TM/MT environment", in: Bernadette Sharp et al. (eds.) *Proceedings of the 8th international NLPCS workshop – Special theme: Human-machine interaction in translation*. Copenhagen Studies in Language 41. Frederiksberg: Samfundslitteratur. 107-118.



- Uswak, V. 2014. "Einsatz der Maschinellen Übersetzung im Übersetzungsprozess in Unternehmen", in: Jan Mugele et al. (eds.). *Tagungsband der Nachwuchswissenschaftlerkonferenz*. Magdeburg, 24 April 2014. 70-73.
- Vázquez, Lucía Morado et al. 2013. "Comparing forum data post-editing performance using translation memory and Machine Translation output: a pilot study", in: Sharon O'Brien, Michel Simard and Lucia Specia. (eds.) *Proceedings of the MT summit XIV workshop on post-editing technology and practice*. Nice, France, September 2 2013: EAMT, 249-256.
- Watkins, Gareth. 2012. *Translation Tools and Technologies in the Welsh Language Context*. Unpublished PhD Thesis, Swansea University.
- Welsh Government and Welsh Language Commissioner. 2015. *National Survey for Wales, 2013-2014: Welsh Language Use Survey*. Cardiff: Welsh Government and Welsh Language Commissioner.
- Welsh Government. 2012. *A Living Language: A Language for Living – Welsh Language Strategy 2012-2017*. Cardiff: Welsh Government.
- Welsh Government. 2014. *A Living Language: A Language for Living – Moving Forward. Policy Statement*. Cardiff: Welsh Government.
- Welsh Government. 2017. *Cymraeg 2050 – A Million Welsh Speakers*. Cardiff: Welsh Government.
- Welsh Language Commissioner. 2012. *Advice Document: Bilingual Drafting, Translation and Interpreting*. Cardiff: Welsh Language Commissioner.
- Welsh Language Commissioner. 2015. *Annual 5-year Report*. Cardiff: Welsh Language Commissioner.
- Yamada, Masaru. 2011. "The effect of translation memory databases on productivity", in: Anthony Pym (ed.) *Translation research projects 3*. Tarragona: Intercultural Studies Group, 63-73.
- Zaretskaya, Anna. 2015. "Integration of Machine Translation in CAT tools: State of the art, evaluation and user attitudes", *Skase Journal of Translation and Interpretation* 8(1): 76-89.
- Zetzsche, Jost. 2007. "Translation Memory: State of the technology", *Multilingual Computing* 8(6), 34-38.
- Zhechev, Ventsislav. 2014. "Analysing the post-editing of Machine Translation at Autodesk", in: Sharon O'Brien et al. (eds.) *Post-editing of Machine Translation: Processes and applications*. Newcastle: Cambridge Scholars Publishing, 2-24.

## LIST OF TABLES

- Table 1: Details of the amount processed by each translator
- Table 2: Pairwise Comparisons for Match Type from the Repeated Measures ANOVA
- Table 3: Time savings as a percentage
- Table 4: Results of Wilcoxon Signed Rank Tests

## LIST OF FIGURES

- Figure 1: Match Type and Time (seconds)
- Figure 2: Median Processing Time in seconds per Welsh translator
- Figure 3: Quality Score according to Match Type